

The Finite Sample Breakdown Point of PCS

Eric Schmitt^a, Viktoria Öllerer^b, Kaveh Vakili^c

^aCorresponding author. *Afdeling Statistiek, Celestijnenlaan 200b - bus 2400, 3001 Leuven. Tel +32 16 37 23 40.*

^b*Faculty of Business and Economics, ORSTAT, KU Leuven, Belgium.*

^c*Afdeling Statistiek, Celestijnenlaan 200b - bus 2400, 3001 Leuven.*

Abstract

The Projection Congruent Subset (PCS) is new method for finding multivariate outliers. PCS returns an outlyingness index which can be used to construct affine equivariant estimates of multivariate location and scatter. In this note, we derive the finite sample breakdown point of these estimators.

Keywords: breakdown point, robust estimation, multivariate statistics.

1. Introduction

Outliers are observations that depart from the pattern of the majority of the data. Identifying outliers is a major concern in data analysis because a few outliers, if left unchecked, can exert a disproportionate pull on the fitted parameters of any statistical model, preventing the analyst from uncovering the main structure in the data.

To measure the robustness of an estimator to the presence of outliers in the data, Donoho (1982) introduced the notion of finite sample breakdown point. Given a sample and an estimator, this is the smallest number of observations that need to be replaced by outliers to cause the fit to be arbitrarily far from the values it would have had on the original sample. Remarkably, the finite sample breakdown point of an estimator can be derived without recourse to concepts of chance or randomness using geometrical features of

Email addresses: eric.schmitt@wis.kuleuven.be (Eric Schmitt), viktoria.oellerer@kuleuven.be (Viktoria Öllerer), kaveh.vakili@wis.kuleuven.be (Kaveh Vakili)

a sample alone (Donoho, 1982). Recently, Vakili and Schmitt (2014) introduced the Projection Congruent Subset (PCS) method. PCS computes an outlyingness index, as well as estimates of location and scatter derived from it. The objective of this paper is to establish the finite sample breakdown of these estimators and show that they are maximal.

Formally, we begin from the situation whereby the data matrix \mathbf{X} , is a collection of n so called *genuine* observations drawn from a p -variate model F with $p > 1$. However, we do not observe \mathbf{X} but an $n \times p$ (potentially) corrupted data set \mathbf{X}^ε that consists of $g < n$ observations from \mathbf{X} and $c = n - g$ arbitrary values, with $\varepsilon = c/n$, denoting the (unknown) rate of contamination.

Historically, the goal of many robust estimators has been to achieve high breakdown while obtaining reasonable efficiency. PCS belongs to a small group of robust estimators that have been designed to also have low bias (see Maronna et al. (1992), Androver and Yohai (2002) and Androver and Yohai (2010)). In the context of robust estimation, a low bias estimator reliably finds a fit close to the one it would have found without the outliers, when $c \leq n - h$ with $h = \lceil (n + p + 1)/2 \rceil$. To the best of our knowledge, PCS is the first member of this group of estimators to be supported by a fast and affine equivariant algorithm (FastPCS) enabling its use by practitioners.

The rest of this paper unfolds as follows. In Section 2, we detail the PCS estimator. In Section 3, we formally detail the concept of finite sample breakdown point of an estimator and establish the notational conventions we will use throughout. Finally, in Section 4, we prove the finite sample breakdown point of PCS.

2. The PCS criterion

Consider a potentially contaminated data set \mathbf{X} of n vectors $\mathbf{x}_i \in \mathbb{R}^p$, with $n > p + 1 > 2$. Given all $M = \binom{n}{h}$ possible h -subsets $\{H^m\}_{m=1}^M$, PCS looks for the one that is most *congruent* along many univariate projections. Formally, given an h -subset H^m , we denote $B(H^m)$ the set of all vectors normal to hyperplanes spanning a p -subset of H^m . More precisely, all directions $\mathbf{a} \in B(H^m)$ define hyperplanes $\{\mathbf{x} \in \mathbb{R}^p : \mathbf{x}'\mathbf{a} = 1\}$ that contain p observations of H^m . For $\mathbf{a} \in B(H^m)$ and $\mathbf{x}_i \in \mathbf{X}$, we can compute the squared orthogonal distance, d_i^2 , of \mathbf{x}_i to the hyperplane defined by \mathbf{a} as

$$d_i^2(\mathbf{a}) = \frac{(\mathbf{a}'\mathbf{x}_i - 1)^2}{\|\mathbf{a}\|^2}. \quad (2.1)$$

The set of the h observations with smallest $d_i^2(\mathbf{a})$ is then defined as

$$H^{\mathbf{a}} = \{i : d_i^2(\mathbf{a}) \leq d_{(h)}^2(\mathbf{a})\}, \quad (2.2)$$

where $d_{(h)}$ denotes the h th-order statistic of a vector \mathbf{d} .

We begin by considering the case in which $d_{(h)}^2(\mathbf{a}) > 0$. For a given subset H^m and direction \mathbf{a} we define the *incongruence index* of H^m along \mathbf{a} as

$$I(H^m, \mathbf{a}) := \log \left(\frac{\text{ave}_{i \in H^m} d_i^2(\mathbf{a})}{\text{ave}_{i \in H^{\mathbf{a}}} d_i^2(\mathbf{a})} \right) \quad (2.3)$$

with the conventions that $\log(0/0) := 0$. This index is always positive and will be smaller the more members of H^m correspond with, or are similar to, the members of $H^{\mathbf{a}}$. To remove the dependency of Equation (2.3) on \mathbf{a} , we measure the incongruence of H^m by considering the average over many directions $\mathbf{a} \in B(H^m)$ as

$$I(H^m) := \text{ave}_{\mathbf{a} \in B(H^m)} I(H^m, \mathbf{a}). \quad (2.4)$$

The optimal h -subset, H^* , is the one satisfying the PCS criterion:

$$H^* = \underset{\{H^m\}_{m=1}^M}{\operatorname{argmin}} I(H^m).$$

Then, the *PCS estimators of location and scatter* are the sample mean and covariance of the observations with indexes in H^* :

$$(\mathbf{t}^*(\mathbf{X}), \mathbf{S}^*(\mathbf{X})) = \left(\text{ave}_{i \in H^*} \mathbf{x}_i, \text{cov}_{i \in H^*} \mathbf{x}_i \right).$$

Finally, we have to account for the special case where $d_{(h)}^2(\mathbf{a}) = 0$. In this case, we enlarge H^* to be the subset of all observations lying on \mathbf{a} . More precisely, if $\exists \mathbf{a}^* \in B(H^*) : |\{i : d_i^2(\mathbf{a}^*) = 0\}| \geq h$, then $H^* = \{i : d_i^2(\mathbf{a}^*) = 0\}$.

2.1. Illustrative Example

To give additional insight into PCS and the characterization of a cloud of point in terms of congruence, we provide the following example. Figure 1 depicts a data set \mathbf{X}^ε of 100 observations, 30 of which come from a cluster

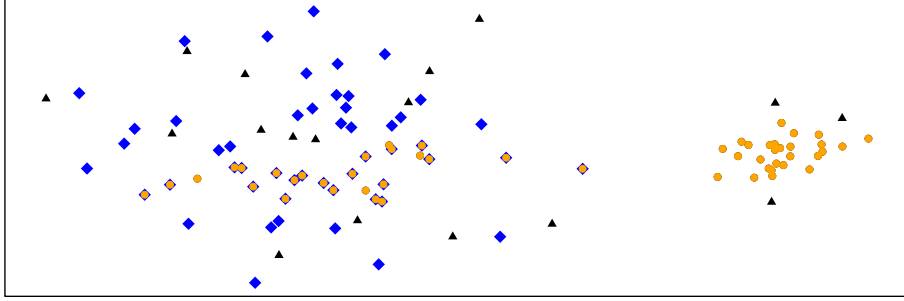


Figure 1: Bivariate data example. The members of H^1 (H^2) are depicted as dark blue diamonds (light orange circles).

of outliers on the right. For this data set, we draw two h -subsets of 52 observations each.

Subset H^1 (dark blue diamonds) contains only genuine observations, while subset H^2 (light orange circles) contains 27 outliers and 25 genuine observations. Finally, the 17 observations belonging to neither h -subset are depicted as black triangles. For illustration's sake, we selected the members of H^2 so that their covariance has smaller determinant than *any* h -subsets formed of genuine observations. Consequently, robust methods based on a characterization of h -subsets in terms of density alone will always prefer the contaminated subset H^2 over any uncontaminated h -subset (and in particular H^1).

The outlierness index computed by PCS differs from that of other robust estimators in two important ways. First, in PCS, the data is projected onto directions given by p points drawn from the members of a given subset, H^m , rather than indiscriminately from the entire data set. This choice is motivated by the fact that when ε and/or p are high, the vast majority of random p -subsets of $\{1, \dots, n\}$ will be contaminated. If the outliers are concentrated, this yields directions almost parallel to each other. In contrast, for an uncontaminated H^m , our sampling strategy always ensures a wider spread of directions and this yields better results.

The second feature of PCS is that the congruence index used to characterize an h -subset depends on all the data points in the sample. We will illustrate this by considering all $\binom{52}{2} = 1326$ members of $B(H^1)$. For each, we compute the corresponding value of $I(H^1, \mathbf{a})$. Then, we sort these and plot them in Figure 2. We do the same for H^2 . We note in passing that $I(H^1) < I(H^2)$.

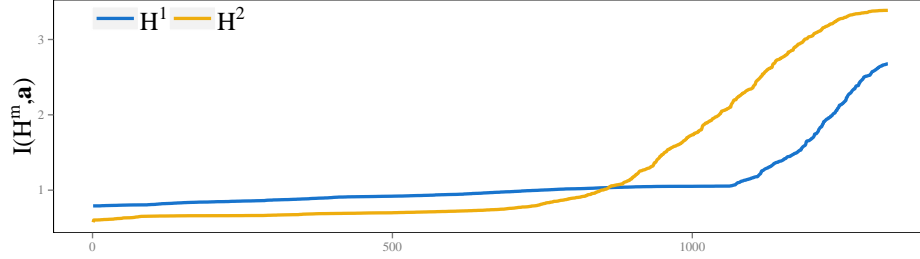


Figure 2: Sorted values of $I(H^m, \mathbf{a})$ for H^1 and H^2 , shown as dark blue (light orange) lines.

Consider now in particular the values of the I -index corresponding to H^1 and starting at around 1050 on the horizontal axis of Figure 2. These higher values of $I(H^1, \mathbf{a})$ correspond to members of $B(H^1)$ that are aligned with the vertical axis (i.e. they correspond to horizontal hyperplanes), and are much larger than the remaining values of $I(H^1, \mathbf{a})$. This is because, for the data configuration shown in Figure 1, the outliers do not stand out from the good data in terms of their orthogonal distances to hyperplanes defined by the vertical directions. As a result, this causes many outliers to enter the sets $H^{\mathbf{a}}$ and this deflates the values of the $\text{ave}_{i \in H^{\mathbf{a}}} d_i^2(\mathbf{a})$ corresponding to these directions. Since the set H^1 is fixed, there is no corresponding effect on $\text{ave}_{i \in H^1} d_i^2(\mathbf{a})$ so that the outliers will influence the values of $I(H^1, \mathbf{a})$ for some directions \mathbf{a} , even though H^1 itself is uncontaminated. This apparent weakness is an inherent feature of PCS. In the remainder of this note, we prove the following counter-intuitive fact: outliers influence the value of $I(H^m)$, even when H^m is free of outliers, yet, so long as there are fewer than $n - h$ of them, their influence on the PCS fit will always remain bounded. In other words, breakdown only occurs if $c > n - h$ (see Section 4).

3. Finite sample breakdown point

To lighten notation and without loss of generality, we arrange the observed data matrix $\mathbf{X}^\varepsilon = ((\mathbf{X}^g)', (\mathbf{X}^c)')'$ with rows $\{\mathbf{x}_i^\varepsilon\}_{i=1}^n$ so that the $\varepsilon\%$ of contaminated observations \mathbf{X}^c are in the last c rows and the uncontaminated observations \mathbf{X}^g in the first g rows. Then, \mathcal{X}^ε will refer to the set of all corrupted data sets \mathbf{X}^ε and \mathcal{H} is the set of all h -subsets of $\{1, \dots, n\}$, $\mathcal{H}^c = \{H \in \mathcal{H} : H \cap \{g + 1, \dots, n\} \neq \emptyset\}$ the set of all

h -subsets of $\{1, \dots, n\}$ with at least one contaminated observation, and $\mathcal{H}^g = \{H \in \mathcal{H} : H \cap \{g+1, \dots, n\} = \emptyset\}$ the set of all uncontaminated h -subsets of $\{1, \dots, n\}$.

The following assumptions (as per, for example Tyler (1994)) all pertain to the original, uncontaminated, data set \mathbf{X} . In the first part of this note, we will consider the case whereby the point cloud formed by $\mathbf{X}^g = \{\mathbf{x}_i^g\}_{i=1}^g$ lies in *general position* in \mathbb{R}^p . The following definition of *general position* is adapted from Rousseeuw and Leroy (1987):

DEFINITION 1: *General position in \mathbb{R}^p .* \mathbf{X} is in general position in \mathbb{R}^p if no more than p -points of \mathbf{X} lie in any $(p-1)$ -dimensional affine subspace. For p -dimensional data, this means that there are no more than p points of \mathbf{X} on any hyperplane, so that any $p+1$ points of \mathbf{X} always determine a p -simplex with non-zero determinant.

Throughout, we will also assume that

$$\sup_{\mathbf{a} \in \mathbb{R}^p_{\neq 0}} \max \left(\frac{\mathbf{X}'\mathbf{a}}{\|\mathbf{a}\|} \right)^2 \text{ is bounded}$$

and that the genuine observations contain no duplicates:

$$\|\mathbf{x}_i - \mathbf{x}_j\| > 0 \forall 1 \leq i < j \leq n.$$

For any h -subset $H^m \in \mathcal{H}$ and \mathbf{X}^ε , we will denote the sample mean and covariance of the observations with indexes in H^m as

$$(\mathbf{t}^m(\mathbf{X}^\varepsilon), \mathbf{S}^m(\mathbf{X}^\varepsilon)) = \left(\text{ave}_{i \in H^m} \mathbf{x}_i^\varepsilon, \text{cov}_{i \in H^m} \mathbf{x}_i^\varepsilon \right).$$

Then, given $\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon$, and an affine equivariant estimator of location \mathbf{t} , we define the bias of \mathbf{t} at \mathbf{X} as

$$\text{bias}(\mathbf{t}, \mathbf{X}, \varepsilon) = \sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \|\mathbf{t}(\mathbf{X}^\varepsilon) - \mathbf{t}(\mathbf{X})\|.$$

Furthermore, given $\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon$, genuine data \mathbf{X} and an affine equivariant estimator of scatter \mathbf{S} with $\mathbf{S}(\mathbf{X})$ positive definite (denoted from now on by $\mathbf{S}(\mathbf{X}) \succ 0$), we define the bias of \mathbf{S} at \mathbf{X} as

$$\text{bias}(\mathbf{S}, \mathbf{X}, \varepsilon) = \sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \frac{\lambda_1(\mathbf{Q}^\varepsilon)}{\lambda_p(\mathbf{Q}^\varepsilon)},$$

where $\mathbf{Q}^\varepsilon = (\mathbf{S}(\mathbf{X})^{-1/2} \mathbf{S}(\mathbf{X}^\varepsilon) \mathbf{S}(\mathbf{X})^{-1/2}) \succ 0$ and $\lambda_1(\mathbf{Q}^\varepsilon)$ ($\lambda_p(\mathbf{Q}^\varepsilon)$) denotes the largest (smallest) eigenvalue of a matrix \mathbf{Q}^ε . Since PCS is affine equivariant (see Appendix 1), w.l.o.g., we can set $\mathbf{t}(\mathbf{X}) = \mathbf{0}$ so that the expression of bias reduces to

$$\text{bias}(\mathbf{t}, \mathbf{X}, \varepsilon) = \sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \|\mathbf{t}(\mathbf{X}^\varepsilon)\|.$$

Furthermore, if that data is in general position and $\mathbf{S}(\mathbf{X})$ is affine equivariant then we can w.l.o.g. set $\mathbf{S}(\mathbf{X}) = \mathbf{I}_p$ (\mathbf{I}_p is the rank p identity matrix) so that the expression of the bias reduces to

$$\text{bias}(\mathbf{S}, \mathbf{X}, \varepsilon) = \sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \frac{\lambda_1(\mathbf{S}(\mathbf{X}^\varepsilon))}{\lambda_p(\mathbf{S}(\mathbf{X}^\varepsilon))}.$$

The finite sample breakdown points ε_n^* (Maronna et al. , 2006) of \mathbf{t} and \mathbf{S} are then defined as

$$\begin{aligned} \varepsilon_n^*(\mathbf{t}, \mathbf{X}) &= \min_{1 \leq c \leq n} \left\{ \varepsilon = \frac{c}{n} : \text{bias}(\mathbf{t}, \mathbf{X}, \varepsilon) = \infty \right\} \\ \varepsilon_n^*(\mathbf{S}, \mathbf{X}) &= \min_{1 \leq c \leq n} \left\{ \varepsilon = \frac{c}{n} : \text{bias}(\mathbf{S}, \mathbf{X}, \varepsilon) = \infty \right\}. \end{aligned}$$

Finally, for point clouds \mathbf{X} lying in general position in \mathbb{R}^p , Davies (1987) gives a strict upper bound for the finite sample breakdown point for any affine equivariant location and scatter statistics (\mathbf{t}, \mathbf{S}) , namely:

$$\begin{aligned} \varepsilon_n^*(\mathbf{t}, \mathbf{X}) &\leq (n - h + 1)/n \\ \varepsilon_n^*(\mathbf{S}, \mathbf{X}) &\leq (n - h + 1)/n \end{aligned} \tag{3.1}$$

4. Finite sample breakdown point of PCS

To establish the breakdown point of \mathbf{S}^* , we first introduce two lemmas describing properties of the I -index. Both deal with the case where \mathbf{X} lies in general position in \mathbb{R}^p . Then, we discuss the case where \mathbf{X} does not lie in general position.

In the first lemma, we show that the incongruence index of a clean h -subset is bounded.

Lemma 1. *Let $c \leq n - h$ and \mathbf{X} lies in general position in \mathbb{R}^p . Then*

$$\sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \max_{H^g \in \mathcal{H}^g} \max_{\mathbf{a} \in B(H^g)} I(H^g, \mathbf{a}) \leq k(\mathbf{X}) \tag{4.1}$$

for any fixed, positive scalar $k(\mathbf{X})$ not depending on the outliers.

Proof. Consider first the numerator of $I(H^g, \mathbf{a})$. For a fixed $H^g \in \mathcal{H}^g$, we can find for each $\mathbf{a} \in B(H^g)$, the p observations of \mathbf{X}^g that lie furthest away from the hyperplane defined by \mathbf{a} . The average of their distances (as given by Equation (2.1)) to the hyperplane \mathbf{a} is finite and constitutes an upper bound on the average distance of any p observations of \mathbf{X}^g to the hyperplane \mathbf{a} . As we have at most $\binom{h}{p}$ different directions $\mathbf{a} \in B(H^g)$ and only $\binom{n-c}{h}$ uncontaminated subsets $H^g \in \mathcal{H}^g$, the upper bound of the average distances stays finite

$$\max_{H^g \in \mathcal{H}^g} \max_{\mathbf{a} \in B(H^g)} \text{ave}_{i \in H^g} d_i^2(\mathbf{a}) \leq U(\mathbf{X})$$

for any positive, fixed, finite scalar $U(\mathbf{X})$ not depending on the outliers. Since the contaminated observations have no influence on the distance $d_i^2(\mathbf{a})$ with $i \in H^g(\mathbf{a})$ for $\mathbf{a} \in B(H^g)$, we can say that

$$\sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \max_{H^g \in \mathcal{H}^g} \max_{\mathbf{a} \in B(H^g)} \text{ave}_{i \in H^g} d_i^2(\mathbf{a}) \leq U(\mathbf{X}). \quad (4.2)$$

Consider now the denominator of $I(H^g, \mathbf{a})$. For any $H^g \in \mathcal{H}^g$ and $\mathbf{a} \in B(H^g)$, let $H^\mathbf{a}$ denote the subset that consists of the indexes of the h observations of the observed data matrix \mathbf{X}^ε that lie closest to the hyperplane spanned by \mathbf{a} . As $c \leq n - h$ and $h = \lceil (n + p + 1)/2 \rceil$, $H^\mathbf{a}$ contains at least $p + 1$ uncontaminated observations. In total, when $H^g \in \mathcal{H}^g$ is not fixed, there are at most $\binom{n-c}{p}$ different directions \mathbf{a} defined by a $H^g \in \mathcal{H}^g$. For any \mathbf{a} , the smallest value of $\text{ave}_{i \in H^\mathbf{a}} d_i^2(\mathbf{a})$ is attained if the contaminated observations of $H^\mathbf{a}$ achieve $d_i^2(\mathbf{a}) = 0$. As the uncontaminated observations lie in general position, we know that the $p + 1$ uncontaminated observations in $H^\mathbf{a}$ cannot lie within the same p -dimensional subspace, i.e.

$$\exists i \in H^\mathbf{a} : d_i^2(\mathbf{a}) > 0.$$

As the number of uncontaminated observations is fixed, we have that

$$\min_{H^g \in \mathcal{H}^g} \min_{\mathbf{a} \in B(H^g)} \text{ave}_{i \in H^\mathbf{a}} d_i^2(\mathbf{a}) \geq l(\mathbf{X}) > 0 \quad (4.3)$$

for any fixed positive scalar $l(\mathbf{X})$ not depending on the outliers. This inequality holds even if the outliers have the smallest average distance that is possible (i.e. when $\mathbf{a} : d_i^2(\mathbf{a}) = 0$ for the contaminated observations). Thus, Inequality (4.3) holds for any ε -contaminated data set \mathbf{X}^ε yielding

$$\inf_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \min_{H^g \in \mathcal{H}^g} \min_{\mathbf{a} \in B(H^g)} \text{ave}_{i \in H^\mathbf{a}} d_i^2(\mathbf{a}) \geq l(\mathbf{X}) > 0. \quad (4.4)$$

Using Equation (2.3) and the Inequalities (4.2) and (4.4), we get (4.1). \square

The second lemma shows the unboundedness of the incongruence index of contaminated subsets.

Lemma 2. *Let $c \leq n - h$ and assume that \mathbf{X} lies in general position in \mathbb{R}^p . Take a fixed h -subset $H^c \in \mathcal{H}^c$. Then*

$$\forall U_1 > 0 : \exists \mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon : I(H^c, \mathbf{a}) > U_1 \quad (4.5)$$

for at least one $\mathbf{a} \in B(H^c)$. In other words, for a given set of indexes H^c , there exists a data set \mathbf{X}^ε with contaminated observations with indexes in H^c such that $I(H^c, \mathbf{a})$ is unbounded.

Proof. Consider first the numerator of $I(H^c, \mathbf{a})$. For a fixed $H^c \in \mathcal{H}^c$, denote $G^+ = \{G \cap H^c\}$. Since $c \leq n - h$, as already mentioned in Lemma 1 above, any h -subset contains at least $p + 1$ uncontaminated observations, i.e. $|G^+| \geq p + 1$. Let $B^+(H^c) \subseteq B(H^c)$ be the set of all directions defining a hyperplane spanned by a p -subset of G^+ . $|G^+| \geq p + 1$ yields $|B^+(H^c)| \geq p + 1$. As the uncontaminated observations $G \supseteq G^+$ lie in general position, the members of $B^+(H^c)$ are, by definition, linearly independent. As a result, the outliers can belong to (at most) the subspace spanned by p uncontaminated observations. Hence, for every $U_2 > 0$, there exists at least one member \mathbf{a}_+^c of $B^+(H^c)$, at least one $i \in H^c$ and at least one $\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon$ such that

$$d_i^2(\mathbf{a}_+^c) > U_2. \quad (4.6)$$

Consider now the denominator of $I(H^c, \mathbf{a})$: Since the members of $B^+(H^c)$ all pass through members of \mathbf{X}^g only, we have that

$$d_{(h)}^2(\mathbf{a}_+^c) \leq U_3 \leq \max_{i \leq h} d_i^2(\mathbf{a}_+^c). \quad (4.7)$$

Using Equation (2.3), and Inequalities (4.6) and (4.7), we get (4.5). \square

With Lemmas 1 and 2, we are now able to derive the finite sample breakdown point of the PCS of \mathbf{S}^* and \mathbf{t}^* .

Theorem 1. *For $n > p + 1 > 2$ and \mathbf{X} in general position, the finite sample breakdown point of \mathbf{S}^* is*

$$\varepsilon_n^*(\mathbf{S}, \mathbf{X}) = \frac{n - h + 1}{n}.$$

Proof. Consider first the situation where $c \leq n - h$. Then any h -subset H^m of \mathbf{X}^ε contains at least $p+1$ members of G . In particular, for the chosen h -subset H^* , denote $G^* = \{H^* \cap \{1, \dots, g\}\}$ with $|G^*| \geq p+1$. The members of G^* are in general position so that $\text{ave}_{i \in G^*}(\mathbf{x}_i^\varepsilon - \mathbf{t})(\mathbf{x}_i^\varepsilon - \mathbf{t})' \succ 0$ for any $\mathbf{t} \in \mathbb{R}^p$. But $\mathbf{S}^*(\mathbf{X}^\varepsilon) = \text{ave}_{i \in H^*}(\mathbf{x}_i^\varepsilon - \mathbf{t}^*)(\mathbf{x}_i^\varepsilon - \mathbf{t}^*)'$ and $\text{ave}_{i \in H^* \setminus G^*}(\mathbf{x}_i^\varepsilon - \mathbf{t}^*)(\mathbf{x}_i^\varepsilon - \mathbf{t}^*)' \succeq 0$ so that $\mathbf{S}^*(\mathbf{X}^\varepsilon) \succ 0$ (Seber, 2008, 10.58) which implies that $\sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \lambda_p(\mathbf{S}^*(\mathbf{X}^\varepsilon)) > 0$.

Thus for breakdown to occur, the numerator of Equation (3.1), $\lambda_1(\mathbf{S}^*(\mathbf{X}^\varepsilon))$, must become unbounded. Now, suppose that $\mathbf{S}^*(\mathbf{X}^\varepsilon)$ breaks down. This means that for any $U_4 > \max_{i=1}^n \|\mathbf{x}_i\|^2$,

$$\sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \lambda_1(\mathbf{S}^*(\mathbf{X}^\varepsilon)) > U_4. \quad (4.8)$$

We will show that this leads to a contradiction. In Appendix 2 we show that

$$\lambda_1(\mathbf{S}^*(\mathbf{X}^\varepsilon)) \leq \max_{i \in H^*} \|\mathbf{x}_i^\varepsilon\|^2. \quad (4.9)$$

By Equations (4.8) and (4.9), it follows that $\sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} \max_{i \in H^*} \|\mathbf{x}_i^\varepsilon\|^2 > U_4$. Then, by Lemma 2 we have that $\sup_{\mathbf{X}^\varepsilon \in \mathcal{X}^\varepsilon} I(H^*) > U_1/K$ with $K = \binom{h}{p}$, the number of all directions $\mathbf{a} \in B(H^*)$. In particular, this is also true for $U_1 > k(\mathbf{X})$, and by Lemma 1, $k(\mathbf{X}) \geq I(H^g)$, implying that $I(H^*) > I(H^g) \forall H^g \in \mathcal{H}^g$, which is a contradiction to the definition of H^* . Since PCS is affine and shift equivariant, when $c > n - h$, we have by Equation (3.1) that $\mathbf{S}^*(\mathbf{X}^\varepsilon)$ breaks down. \square

Equation (3.1) and Theorem 1 show that the breakdown point of \mathbf{S}^* is maximal. The following theorem shows that the breakdown point of \mathbf{t}^* is also maximal.

Theorem 2. *For $n > p + 1 > 2$ and \mathbf{X} in general position, the finite sample breakdown point of \mathbf{t}^* is*

$$\varepsilon_n^*(\mathbf{t}, \mathbf{X}) = \frac{n - h + 1}{n}.$$

Proof. Consider first the situation where $c \leq n - h$. In Theorem 1, we showed that under this condition $\max_{i \in H^*} \|\mathbf{x}_i^\varepsilon\| < \infty$. Denote \mathbf{x}_u^ε , where $u =$

$\operatorname{argmax}_{i \in H^*} \|\mathbf{x}_i^\varepsilon\|$. Then, we have that $\mathbf{t}^*(\mathbf{X}^\varepsilon)$ does not break down since, by homogeneity of the norm and the triangle inequality,

$$\|\mathbf{t}^*(\mathbf{X}^\varepsilon)\| \leq \frac{1}{h} \sum_{i \in H^*} \|\mathbf{x}_i^\varepsilon\| \leq \frac{1}{h} \sum_{i=1}^h \|\mathbf{x}_u^\varepsilon\| = \|\mathbf{x}_u^\varepsilon\| < \infty.$$

For the case of $c > n - h$, Equation (3.1) and affine equivariance imply that $\mathbf{t}^*(\mathbf{X}^\varepsilon)$ breaks down. \square

We now relax the assumption that the members of \mathbf{X}^g lie in general position in \mathbb{R}^p and substitute it by the weaker condition that they all lie in general position on a common subspace in \mathbb{R}^q for some $q < p$. Then PCS has the so-called exact fit property. Recall that \mathbf{a} are hyperplanes defined by p points drawn from an h -subset $H \in \mathcal{H}$. If there are at least h points lying on a subspace, then there exists an h -subset of points from this subspace. Let \tilde{H} be this subset. Then, for any $\mathbf{a}^+ \in B(\tilde{H})$, both the numerator and denominator of Equation (2.3) equal zero and so $I(\tilde{H}) = 0$. Thus, we have without loss of generality that $H^* = \{i : d_i^2(\mathbf{a}^+) = 0\}$. In summary this means that if h or more observations lie exactly on a subspace, the fit given by the observations in H^* will coincide with this subspace, which is the definition of the so-called *exact fit* property. Of course, since $|H^*| \geq h$, H^* may contain outliers. Given H^* , one may proceed with the much simpler task of identifying the at most $|H^*| - h$ outliers in this smaller set of observations on a rank q subspace spanned by the members of H^* .

Appendix 1: Proof of affine equivariance of $(\mathbf{t}^*(\mathbf{X}^\varepsilon), \mathbf{S}^*(\mathbf{X}^\varepsilon))$

Recall that a location vector $\mathbf{t}(\mathbf{X}^\varepsilon)$ and a scatter matrix $\mathbf{S}(\mathbf{X}^\varepsilon)$ are affine equivariant if for any non-singular $p \times p$ matrices \mathbf{B} and p -vector \mathbf{b} it holds that:

$$\begin{aligned} \mathbf{t}(\mathbf{B}\mathbf{X}^\varepsilon + \mathbf{1}_n \mathbf{b}') &= \mathbf{B}\mathbf{t}(\mathbf{X}^\varepsilon) + \mathbf{b} \\ \mathbf{S}(\mathbf{B}\mathbf{X}^\varepsilon + \mathbf{1}_n \mathbf{b}') &= \mathbf{B}\mathbf{S}(\mathbf{X}^\varepsilon)\mathbf{B}'. \end{aligned}$$

Consider now affine transformations of \mathbf{X}^ε :

$$\mathbf{y}_i^\varepsilon = \mathbf{B}\mathbf{x}_i^\varepsilon + \mathbf{b}', \quad i = 1, \dots, n. \quad (4.10)$$

for any non-singular $p \times p$ matrix \mathbf{B} and p -vector \mathbf{b} . The directions \mathbf{a}_x (\mathbf{a}_y) are orthogonal to hyperplanes through p -subsets of \mathbf{X}^ε (\mathbf{Y}^ε). Since $\|\mathbf{x}_i^g - \mathbf{x}_j^g\| >$

$0 \vee 1 \leq i < j \leq g$, we can disregard all duplicated rows of \mathbf{X}^ε (and their partner duplicates in \mathbf{Y}^ε), so that, w.l.o.g. all p -subsets of \mathbf{X}^ε (\mathbf{Y}^ε) yield a $p \times p$ matrix with unique rows. Let p^0 be any such p -subset of $\{1 : n\}$, and \mathbf{a}_x^0 and \mathbf{a}_y^0 the hyperplanes through $\{\mathbf{x}_i\}_{i \in p^0}$ and $\{\mathbf{y}_i\}_{i \in p^0}$. Since Equation (4.10) describes an affine transformation, it preserves collinearity:

$$\{i : \mathbf{x}'_i \mathbf{a}_x^0 = 1\} = \{i : \mathbf{y}'_i \mathbf{a}_y^0 = 1\}, \quad (4.11)$$

and the ratio of lengths of intervals on univariate projections (Weisstein , 2014, sec. 36):

$$\frac{\text{ave}_{i \in H^m} \|\mathbf{x}'_i \mathbf{a}_x^0 - \text{ave}_{i \in p^0} \mathbf{x}'_i \mathbf{a}_x^0\|}{\text{ave}_{i \in H(\mathbf{a}_x^0)} \|\mathbf{x}'_i \mathbf{a}_x^0 - \text{ave}_{i \in p^0} \mathbf{x}'_i \mathbf{a}_x^0\|} = \frac{\text{ave}_{i \in H^m} \|\mathbf{y}'_i \mathbf{a}_y^0 - \text{ave}_{i \in p^0} \mathbf{y}'_i \mathbf{a}_y^0\|}{\text{ave}_{i \in H(\mathbf{a}_y^0)} \|\mathbf{y}'_i \mathbf{a}_y^0 - \text{ave}_{i \in p^0} \mathbf{y}'_i \mathbf{a}_y^0\|}, \quad (4.12)$$

where for readability we denote $H^{\mathbf{a}}$ as $H(\mathbf{a})$. Equation (4.11) and (4.12) imply

$$I(H^m, \mathbf{a}_x^0) = I(H^m, \mathbf{a}_y^0). \quad (4.13)$$

Equation (4.13) holds for any p -subset of H^m . Therefore, denoting $B_x(H^m)$ all directions perpendicular to hyperplanes through p elements of $\{\mathbf{x}_i^\varepsilon\}_{i \in H^m}$, and $B_y(H^m)$ the same but for $\{\mathbf{y}_i^\varepsilon\}_{i \in H^m}$, it holds that

$$\text{ave}_{\mathbf{a}_x \in B_x(H^m)} I(H^m, \mathbf{a}_x) = \text{ave}_{\mathbf{a}_y \in B_y(H^m)} I(H^m, \mathbf{a}_y), \quad m = 1 \dots, M$$

and in particular for $H^m = H^*$. Since $\#\{H^*\} \geq p + 1$, we have that if the members of H^* lie in G.P. in \mathbb{R}^p ,

$$\begin{aligned} \text{ave}_{i \in H^*} (\mathbf{B} \mathbf{x}_i^\varepsilon + \mathbf{b}') &= \mathbf{B} \text{ave}_{i \in H^*} (\mathbf{x}_i^\varepsilon) \mathbf{b}, \\ \text{cov}_{i \in H^*} (\mathbf{B} \mathbf{x}_i^\varepsilon + \mathbf{b}') &= \mathbf{B} \text{cov}_{i \in H^*} (\mathbf{x}_i^\varepsilon) \mathbf{B}'. \end{aligned}$$

Hence, $(\mathbf{t}^*(\mathbf{X}^\varepsilon), \mathbf{S}^*(\mathbf{X}^\varepsilon))$ are affine equivariant.

Appendix 2: Proof of Equation 4.9

Here, we show that $\lambda_1(\mathbf{S}^*(\mathbf{X}^\varepsilon)) \leq \max_{i \in H^*} \|\mathbf{x}_i^\varepsilon\|^2$. The first eigenvalue of $\mathbf{S}^*(\mathbf{X}^\varepsilon)$ is defined as

$$\lambda_1(\mathbf{S}^*(\mathbf{X}^\varepsilon)) = \text{cov}_{i \in H^*} ((\mathbf{x}_i^\varepsilon)' d)$$

for $d = \operatorname{argmax}_{\|d\|=1} \operatorname{cov}((\mathbf{x}_i^\varepsilon)'d)$. Furthermore,

$$\operatorname{cov}_{i \in H^*}((\mathbf{x}_i^\varepsilon)'d) = \operatorname{ave}_{i \in H^*}(((\mathbf{x}_i^\varepsilon)'d)^2) - (\operatorname{ave}_{i \in H^*}((\mathbf{x}_i^\varepsilon)'d))^2.$$

Hence, we have that

$$\operatorname{cov}_{i \in H^*}((\mathbf{x}_i^\varepsilon)'d) \leq \operatorname{ave}_{i \in H^*}(((\mathbf{x}_i^\varepsilon)'d)^2).$$

We also have that

$$\operatorname{ave}_{i \in H^*}(((\mathbf{x}_i^\varepsilon)'d)^2) \leq \max_{i \in H^*}(((\mathbf{x}_i^\varepsilon)'d)^2) = \max_{i \in H^*} \|(\mathbf{x}_i^\varepsilon)'d\|^2.$$

Using Cauchy-Schwartz,

$$\max_{i \in H^*} \|(\mathbf{x}_i^\varepsilon)'d\|^2 \leq (\max_{i \in H^*} \|d\| \|\mathbf{x}_i^\varepsilon\|)^2,$$

and $\|d\| = 1$. Thus, $\lambda_1(\mathbf{S}^*(\mathbf{X}^\varepsilon)) \leq \max_{i \in H^*} \|\mathbf{x}_i^\varepsilon\|^2$.

5. Acknowledgements

The authors wish to acknowledge the helpful comments from three anonymous referees and the editor for improving this paper.

Viktoria Öllerer would like to acknowledge the support of Research Fund KU Leuven GOA/12/014.

References

- Adrover, J. G. and Yohai, V. J. (2002). Projection estimates of multivariate location. *Annals of Statistics*. Vol. 30, Number 6, 1760–1781.
- Adrover, J. G. and Yohai, V. J. (2010). A new projection estimate for multivariate location with minimax bias. *Journal of Multivariate Analysis*, Vol. 101, Issue 6, 1400–1411.
- Davies, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics*. 15 1269–1292.
- Donoho, D.L. (1982). Breakdown properties of multivariate location estimators Ph.D. Qualifying Paper Harvard University.

- Maronna, R. A. Stahel, W. A. Yohai, V. J. (1992). Bias-robust estimators of multivariate scatter based on projections. *Journal of Multivariate Analysis*, Vol. 42, Issue 1, 141–161.
- Maronna, R. A., Martin R. D. and Yohai V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, New York.
- Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
- Seber, G. A. F. (2008). *Matrix Handbook for Statisticians*. Wiley Series in Probability and Statistics. Wiley, New York.
- Tyler, D.E. (1994). Finite Sample Breakdown Points of Projection Based Multivariate Location and Scatter Statistics. *The Annals of Statistics*, Vol. 22, No. 2, pp. 1024–1044
- Vakili, K. and Schmitt, E. (2014). Finding multivariate outliers with Fast-PCS. *Computational Statistics & Data Analysis*, Vol. 69, 54–66.
- Weisstein, E. W. (2002). *Concise Encyclopedia of Mathematics* (2nd edition). Chapman & Hall CRC.